Self-Checking Monitor for NBTI Due Degradation

M. Omaña D. Rossi N. Bosio C. Metra

DEIS – ARCES, University of Bologna, Italy

{martin.omana, d.rossi, cecilia.metra}@unibo.it; nicolo.bosio@studio.unibo.it

***Abstract*— Performance degradation of integrated circuits due to aging effects, such as Negative Bias Temperature Instability (NBTI), is becoming of great concern for current and future CMOS technology. In this paper we propose a monitor able to detect NBTI due late transitions in the combinational part of a critical data-path. It requires lower area than recently proposed alternative solutions, and a lower or comparable power consumption. Moreover, differently from alternative solutions, our monitor is also self-checking with respect to its possible internal faults, thus avoiding the useless negative impact on system performance and the negative impact on system reliability which could otherwise take place in case of non self-checking sensors, should they get affected by faults.**

Keywords - NBTI degradation; critical data-path; monitor; self-checking circuit

I. INTRODUCTION

The design of reliable circuits is becoming increasingly challenging with scaled CMOS technologies. Aggressive oxide thickness scaling has led to large vertical electric fields in MOSFET devices, making oxide breakdown a critical issue. The high field may also lead to significant threshold voltage shift over time due to Negative-Bias Temperature-Instability (NBTI), creating additional uncertainty in the device behavior [1].

Particularly, NBTI is recognized as the primary parametric failure mechanism in modern ICs [2]. It is characterized by a positive shift in the absolute value of the pMOS threshold voltage, mainly due to the creation of positively charged interface traps when the transistor is biased in strong inversion (that is, when it is off , thus having Vgs = -Vdd) [1, 3]. Due to this phenomenon, the absolute threshold voltage can increase by more than 50mV over ten years [4], resulting in a circuit performance degradation of more than 20% [5].

In case of data-paths, such a threshold voltage increase may lead to a late transition of the signal at the input of an output flip-flop. In case of critical timing paths, such a signal may reach its stable final value later than expected, thus violating the flip-flop set-up time and resulting in a incorrect sampled value.

A straightforward solution could be to increase the clock period by a time interval (usually referred to as guardband [2]) equal to the expected maximum worst case performance degradation due to NBTI over the chip lifetime [2]. However, this would introduce an excessive time margin from the beginning of circuit operation, with a consequent very high (and unnecessary) negative impact on system performance [2]. As an alternative approach, a smaller timing guardband could be adopted, together with circuit failure prediction schemes able to monitor circuit performance degradation, to then adapt accordingly the clock period, throughout the chip lifetime [2, 4]. Compared to the above mentioned worst case scenario, this would allow a lower negative impact on system performance.

More in details, the system would start operating with the highest clock frequency guaranteeing the circuit correct operation. Then, should aging sensors detect late transitions during a pre-defined guardband, the clock period could be locally, or globally, increased (by a proper time interval depending on the estimated amount of circuit aging [2]), in order to avoid the sampling of incorrect data. This approach would allow the system to continue working correctly [1, 6, 4], with consequent reliability increase, at the cost of some performance degradation.

In this regard, it should be considered that transistors' stress can largely vary for different areas of the chip, so that NBTI due late signal transitions will not uniformly occur throughout the chip. Therefore, several aging sensors should be strategically distributed throughout the chip, making the low cost of such sensors be a relevant issue.

Several aging sensors have been proposed up to now to monitor performance degradation due to NBTI (e.g., those in [7, 8, 9, 4, 6]).

In [8, 9], the sensors are implemented by ring-oscillators allowing to identify performance degradation by monitoring possible changes in their oscillation frequency. They feature high measurement accuracy, but require considerable area overhead.

In [7], the effects of NBTI are monitored by sensing a leakage current reduction using I_{DDQ} testing, thus suffering all well known limitations of I_{DDQ} testing for future technologies [10].

More recently, in [4], a compact sensor monitoring critical paths for late transitions during a proper guardband has been

978-1-4244-7793-7/10/$26.00 ©2010 IEEE

proposed. Moreover, in [6], sensors comparing the speed of an inverter stressed with high switching activity to that of a non-stressed inverter have been presented in order to measure the effect of NBTI. The sensors in [4, 6] feature lower power consumption and area overhead compared to previously published sensors.

Based on these considerations, in this paper we propose a new sensor for NBTI degradations that, exploiting the idea presented in [11] to transform late signal transitions into coded/non-coded words for delay and transient faults' detection, allows to monitor performance degradations due to NBTI, at lower area and lower or comparable power consumption, than the previous lowest cost solutions in [4, 6]. Particularly, our circuit monitors the outputs of combinational critical paths during a proper guardband, and produces an *alarm* message in case of late transitions due to NBTI during such a guardband, and a *no alarm* indication otherwise. The basic scheme of our monitor has been introduced in [12].

In addition to its power and/or area advantages over previous solutions, our monitor also features self-checking ability with respect to its possible internal faults (including node stuck-ats, transistor stuck-ons, transistor stuck-opens and resistive bridgings). Particularly, it provides an *alarm* message also in case of faults affecting itself. As usual in case of self-checking circuits [13], our monitor presents two output signals, which are encoded by the two-rail code.

It should be noted that the self-checking ability is a relevant property of our proposed monitor. In fact, as for all other (non self-checking) sensors, it is sufficient that a simple stuck-at fault affects their single output to make: i) the faulty sensor unable to detect late transitions due to NBTI (in case of a stuck-at at the *no alarm* indication), or ii) keeping on (erroneously) indicating the occurrence of late transitions (in case of a stuck-at at the *alarm* indication). In case i), *no alarm* message will be provided in the presence of late transitions due to NBTI, thus making the sensor unable to prevent incorrect data sampling, with consequent negative impact on system reliability, while in case ii), an *alarm* message will continuously be provided (also without any NBTI due late transition) and the system clock frequency will continuously be (uselessly) slowed down, with consequent negative impact on system performance. Self-checking sensors, instead, are able to provide an output *alarm* message in case of faults affecting themselves, included those affecting their outputs. Should their produced *alarm* message persist also after clock frequency decrease, it could be reasonably recognized at the system level as an *alarm* message due to the occurrence of sensor's internal faults, rather than late transitions due to NBTI. Consequently, proper actions could be taken at the system level in order to: a) avoid keeping on decreasing clock frequency in case of persisting *alarm* messages, thus avoiding the useless negative impact on system performance which would otherwise take place in case of non self-checking sensors; b) activate proper recovery techniques to recover from the detected sensor faults, thus guaranteeing the availability of fault-free sensors able to prevent incorrect data sampling, thus avoiding the negative impact on system reliability which would otherwise take place in case of non self-checking sensors.

The rest of this paper is organized as follows. In Section II, we describe our proposed monitor. In Section III, we show some results of the electrical level simulations that we have performed to verify its behavior. In Section IV, we discuss its self-checking ability. In Section V, we evaluate the costs of our monitor and compare them to those of alternate solutions. Finally, some conclusive remarks are drawn in Section VI.

II. PROPOSED MONITOR

Let us consider a generic critical data-path of the kind shown in Fig. 1. The block C_i denotes a combinational circuit, whose worst case propagation delay has been denoted by t_{pd}. FF_i and FF_{i+1} are the input and output flip-flops (FFs). They present a set-up time equal to t_{set}, and their outputs reach a final stable value after a time t_{pcq} from the CK rising (sampling) edge.

For the circuit correct operation, the output of C_i (S_i) should reach its final stable value before the setup time of FF_{i+1}. However, due to NBTI, the performance of C_i can degrade over time, resulting in a late transition of its output, which may become no longer able to satisfy the FF_{i+1} setup time constraint, thus resulting in an incorrect sampled datum.

To avoid this problem, a proper monitor capable of detecting late transitions of Si due to NBTI can be inserted in the data-path, as shown in Fig. 2.

As usual in case of self-checking circuits, our proposed monitor presents two output signals (O_1 and O_2), in order to avoid the discussed above useless negative impact on system performance and hurt of system reliability.

In order to detect NBTI due late transitions, such a circuit should monitor S_i during a proper guardband (T_M). In particular, T_M should be chosen greater than t_{set} (namely, $T_M = \Delta t1 + t_{set}$, with $\Delta t1 > 0$) in order to avoid that NBTI due late transitions of the monitored signal could result in an incorrect sampled datum. In turn, it should be $0 < \Delta t1 < T_{CK} - t_{pcq} - t_{pd} - t_{set}$, in order to avoid the generation of false alarms due to legal signal transitions. Of course, the higher is $\Delta t1$, the maximal is the monitor's ability to detect NBTI due late transitions (thus avoiding possible incorrect data sampling), but also the impact

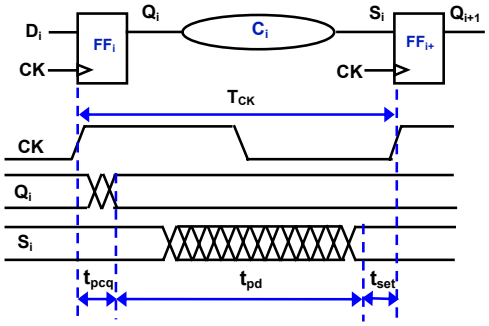

Fig. 1. Schematic representation of the considered data-paths and signals' timing.

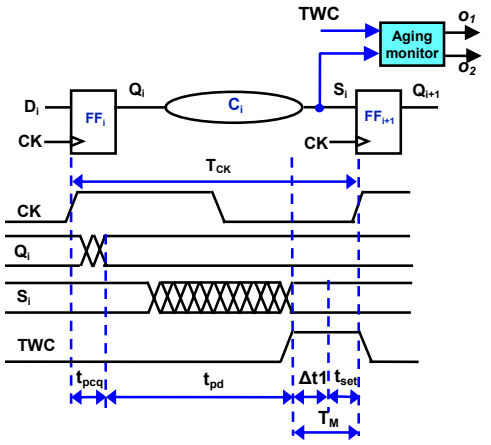

Fig. 2. Monitor insertion within the considered data-path.

on system performance. Therefore, $\Delta t1$ should be properly chosen, in order to trade-off the desired likelihood of system correct operation (thus reliability) and performance. For general purpose applications (with no excessive needs in terms of reliability), $\Delta t1$ can be reasonably chosen based on circuit level estimations of the NBTI degradation in the first 2 weeks [4], or 8 weeks [2] of chip lifetime. As reported in the next section, as an example we have chosen $\Delta t1=30ps$.

To enable our monitor only during T_M, we generate a control signal, denoted as Time Window Control (TWC), which is asserted only during T_M.

The monitor produces two output signals (O_1 and O_2), that are two-rail encoded, thus providing a *no alarm* indication in case of no late transition of Si while TWC = 1. Instead, in case of NBTI due late transitions of Si while TWC = 1, our monitor gives a non two-rail encoded output, thus providing an *alarm* indication.

The internal structure of our proposed monitor is schematically shown in Fig. 3, while the circuit generating TWC is shown in Fig. 4.

Let us start describing in details the internal structure of our monitor. As illustrated in Fig. 3, starting from the monitored signal Si, by means of a proper inverting delay block (e.g., a simple NOT), we generate an additional signal Ci that, together with the monitored signal Si, provides: 1) a word belonging to the two-rail code, that is (O_1,O_2) = (1,0) or (0,1), if Si is stable

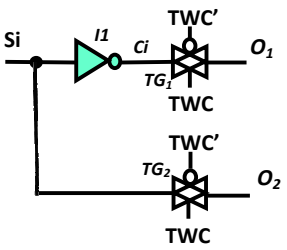

Fig. 3. Internal structure of our monitor.

while TWC = 1; 2) a two-rail non codeword, that is (O_1,O_2) = (0,0) or (1,1), if late transitions of Si occur while TWC=1.

Instead, when TWC = 0, Si transitions are allowed, and the transfer gates disconnect Ci and Si from the monitor outputs O_1 and O_2, respectively.

Let us analyze in more details the behavior of our monitor when TWC = 1. During this time interval, the transfer gates are conductive, and O_1 and O_2 are connected to Ci and Si, respectively. In this case, two conditions may occur, depending on whether the monitored signal is stable (case 1), or presents a late transition (case 2).

In case 1), the logic values (01) or (10), that are present at the inputs of the transfer gates after the last legal signal transition (occurring when TWC=0), are given to (O_1, O_2). Instead, in case 2), a (00) or (11) is produced on (O_1, O_2), whose duration equals the input-output delay of the inverter I1, which can be fixed equal to a chosen value by proper electrical level design.

Therefore, assuming the presence of (01) or (10) on (O_1, O_2) as a *no alarm* message, and that of (00) or (11) as an *alarm* indication, this circuitry can be used to detect on-line late transitions of Si due to NBTI, thus preventing the output FF_{i+1} from sampling an incorrect datum.

As stated before, in case of transitions of Si while TWC=1, our monitor produces an *alarm* indication with a duration equal to the input-output delay of inverter I1. However, should this be needed to allow accomplishment of clock adjustment at the system level, such an *alarm* indication could be maintained for a longer time (particularly till the activation of a proper reset signal), by connecting to the monitor outputs an error indicator (for instance of the kind in [11]).

Now let us describe how the signal TWC can be generated. It can be obtained from the CK signal, by means of a proper pulse generation circuit of the kind illustrated in Fig. 4(a).

As can be seen from the waveforms in Fig 4(b), TWC is equal to 1 only during the time interval $T_{CK}/2$-d, that is when CK=A=0. Accordingly to Fig. 2, $T_{CK}/2$-d is equal to $T_M = \Delta t1 + t_{set}$. Therefore, the delay d can be expressed as: $d = T_{CK}/2 - T_M = T_{CK}/2 - \Delta t1 - t_{set}$. In order to accomplish this equality, the delay element can be implemented by means of a programmable delay element of the kind in [14]. This way, the delay d can be adjusted each time the clock signal is changed, in order to keep T_M equal to the desired value ($\Delta t1 + t_{set}$). The signal TWC' is obtained by simply inverting the signal TWC.

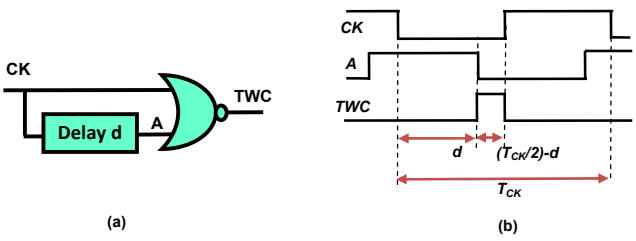

Fig. 4. (a) Circuit generating TWC; (b) timing of its signals.

III. Monitor Implementation and Verification

As an example, we implemented our proposed monitor and pulse generator circuit by means of a standard 45nm CMOS technology, with Vdd = 1V and clock frequency of 3GHz.

The following transistor aspect ratios have been considered: (i) *(W/L) = 1,* for the nMOS of TG1, TG2, the inverter I1 and the NOR gate generating TWC; (ii) *(W/L) = 2,* for the pMOS of TG1, TG2, the inverter I1, and the NOR gate generating TWC.

As an example, we have considered $\Delta t1=30ps$, $t_{set}=15ps$ so that $d= 122ps$. Such a delay d has been obtained implementing the "Delay d" block in Fig. 4(a) by means of 14 minimum-sized, symmetric inverters.

The behavior of our proposed monitor (including the pulse generation circuit) has been verified by means of conventional and Monte Carlo electrical simulations, performed considering statistical variations (with uniform distribution) up to 20% of power supply, oxide thickness, transistor threshold voltage, and electron/hole mobility.

Figure 5 shows the results of the simulations performed under nominal values of electrical parameters, and in case of no late transition of Si occurring while TWC=1. We can observe that, as expected, our sensor gives always complemented outputs O_1 and O_2, thus providing a *no alarm* indication.

Instead, Fig. 6 shows the results of some simulations performed under nominal values of electrical parameters and in case of late transitions of Si occurring while TWC=1. We can observe that, as expected, our monitor produces equal logic values on O_1 and O_2, that is an *alarm* indication. As described above, the duration of such an *alarm* is equal to the input-output delay (τ_{I1}) of the inverter I1 (Fig. 3).

Analogous results have been obtained by means of Monte Carlo simulations, performed considering statistical variations of process parameters.

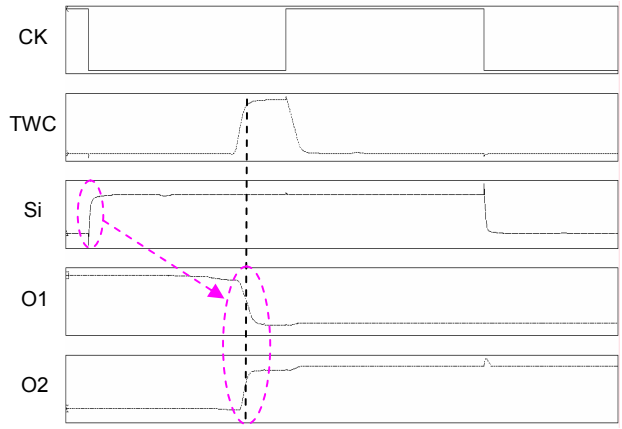

Fig. 5 . Si mulation results o btained f or nominal v alues o f electrical parameters and no late transition of Si occurring while TWC=1.

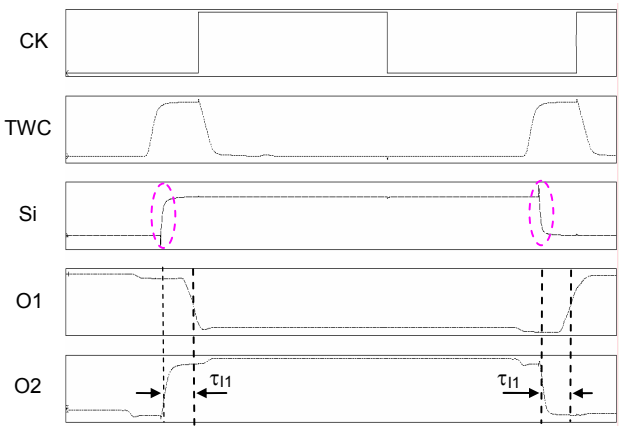

Fig. 6. Si mulation results obtai ned f or nom inal values of electrical parameters and late transitions of Si occurring while TWC=1.

IV. Self-Checking Ability

Similarly to checkers of self-checking circuits (SCCs) [13], we will verify that our monitor is *Totally Self-Checking (TSC)* [13], or *Strongly Code-Disjoint (SCD)* [15], with respect to its internal faults. In particular, we have considered a realistic set of faults (F), composed by all possible node stuck-ats (SAs), transistor stuck-ons (SONs), transistor stuck-opens (SOPs) and resistive bridgings (BFs), with values of connecting resistance (R) in the [0..6kΩ] range [16].

The following fault assumptions [13] are generally considered in case of SCCs: 1) faults occur one at a time; 2) the time elapsing between the occurrence of two following faults is long enough to allow the application of all possible input codewords. In our case, hypothesis 2) can be reasonable changed into the assumption that the time elapsing between the occurrence of two following faults is long enough to allow that Si assumes both a low and a high logic value.

As for SAs 1/0, they may affect: 1) Ci; 2) O_1 or O_2: We have verified that, in both cases 1) and 2), an *alarm* indication (i.e., (O_1,O_2)=(0,0) or (1,1)) is generated before the occurrence of a following fault, and that till *alarm* generation the monitor keeps its ability to detect late transitions of its input signal. Thus, our monitor is *TSC* with respect to them.

Now, let us consider all possible SONs. They might affect a transistor of: 1) TG1 or TG2; 2) inverter I1.

As for SONs of kind 1), we have verified that, as for SAs, our monitor is *TSC* with respect to them.

As for SONs of kind 2), they result in an intermediate voltage value on nodes Ci and O_1 (Fig. 3). Depending on the logic threshold of the downstream logic, such intermediate voltage value may be, or may be not recognized as an *alarm* indication. If an *alarm* indication is recognized, the fault is detected. In case no *alarm* message is recognized, we have verified that the monitor keeps its ability to detect possible late transitions of its input signal during T_M and, should another fault from F occur, our monitor either produces an *alarm* indication, or maintains its ability to detect late transitions of its input signal during T_M. This holds true for all possible sequences of faults in F, so that our monitor is *SCD* with respect to these SONs.

Therefore, our monitor is *TSC* or *SCD* with respect to all possible SONs.

Now, let us consider possible SOPs which might affect a transistor of: 1) TG1 or TG2; 2) the inverter I1.

As for SOPs of kind 1), they make the affected transistor always off. However, TG1 and TG2 are still conductive (during T_M) and, in the absence of late transitions of the monitored signal, no *alarm* indication is produced. However, we have verified that the monitor keeps its ability to detect possible late transitions of its input signal during T_M and, should another fault from F occur, our monitor either produces an *alarm* indication, or maintains its ability to detect late transitions of its input signal during T_M. This holds true for all possible sequences of faults in F, so that our monitor is *SCD* with respect to these SOPs.

As for SOPs of kind 2), considerations similar to those for SA 1/0 affecting node Ci hold true. Therefore, our monitor is *TSC* with respect to this kind of SOPs.

Therefore, our monitor is *TSC* or *SCD* with respect all possible SOPs.

As for all possible BFs, electrical level simulations have been performed by means of HSPICE, considering, for each BF, values of connecting resistance R in the interval [0..6kΩ] [16].

We have verified that our sensor satisfies the *TSC* or the *SCD* property for all possible BFs and values of R, but for two BFs. They are: the BF between the input Si and TWC, and the BF between the input Si and TWC'. Due to these BFs, our monitor may no longer be able to detect late transitions of Si due to NBTI. However, the occurrence of these two BFs could be avoided by properly designing the circuit layout [17].

To summarize, we have found that our proposed monitor satisfies the *TSC,* or the *SCD* property for all faults in F, but for two BFs, whose likelihood could be reduced by means of proper layout design.

V. Cost Evaluation and Comparison

We have evaluated the costs of our proposed monitor in terms of area overhead and power consumption, and we have compared them to those of the recently published aging sensors in [4, 6], which feature the lowest cost over previous solutions.

Electrical level simulations of all compared sensors have been performed considering a standard 45nm CMOS technology, V_{dd} = 1V and a clock frequency of 3GHz. Additionally, all sensors have been implemented assuming the minimum transistor sizes making them work properly. As for the sensor in [6], we have not included the cost of the circuitry required to generate its control signal, since it was not reported. Instead, as for our monitor and that in [4], the cost of the circuit generating the control signal has been included.

As for area overhead, it has been roughly estimated in terms of squares. As for power consumption, we have evaluated the average power consumed by each sensor, considering the case of no late transition of the monitored signal, and monitored signal switching activity of the 25% of the clock frequency. We have also included the static power consumption due to leakage.

The costs of our monitor are reported in Table 1, together with those of the alternate solutions in [4] and [6], and their relative variations over those of our monitor ($\Delta A = 100 \cdot (A_{[4,6]} - A_{our})/A_{our}$; and $\Delta P = 100 \cdot (P_{[4,6]} - P_{our})/P_{our}$).

As can be seen, compared to the sensor in [4], our monitor features significant area reduction, while requiring comparable power consumption. Instead, compared to [6], our monitor requires significantly lower power, as well as lower area overhead (since the cost of implementing the control signal of [6] has not been included in its area evaluation).

TABLE 1. COSTS OF OUR MONITOR AND OF THE SENSOR IN [4] AND IN [6], AS WELL AS RELATIVE COST VARIATIONS.

Considered Sensor	Area (Sq)	ΔA (%)	Power (μW)	ΔP (%)
Our	60	-	12	-
[4]	78	+30%	12.2	+1.7%
[6]	62	+3.3%	15	+25%

These advantages in terms of power consumption or/and area overhead should be added to our monitor provided self-checking ability with respect to its internal faults, thus avoiding the useless negative impact on system performance and the negative impact on system reliability, which (as discussed before) could otherwise take place in case of non self-checking sensors, should they get affected by faults.

As for the resolution of our proposed monitor, analogously to the recently published aging sensors in [4, 6], it is given by the accuracy of the circuit generating TWC.

VI. Conclusions

We have proposed a monitor able to detect NBTI performance degradations in the combinational part of a critical timing data-path. Our monitor requires lower area than recently proposed alternative solutions, and a lower or comparable power consumption. Moreover, our monitor is also self-checking with respect to its possible internal faults, thus avoiding the useless negative impact on system performance and negative impact on system reliability, which could otherwise take place in case of non self-checking sensors, should they get affected by faults.

References

[1] J. Keane, T-H. Kim, C. H. Kim, "An On-Chip NBTI Sensor for Measuring pMOS Threshold Voltage Degradation", IEEE Trans. On Very Large Scale Integration (VLSI) Syst., 2009.

[2] M. Agarwal, B.C. Paul, M. Zhang, S. Mitra, "Circuit Failure Prediction and Its Application to Transistor Aging", in Proc. of IEEE VLSI Test Symp., pp. 277-286, 2007.

[3] V. Huard, M. Denais, "Hole Trapping Effect on Methodology for DC and AC Negative Bias Temperature Instability Measurements in PMOS Transistors", in Proc. of IEEE Int. Rel. Physics Symp., pp 40-45, 2004.

[4] M. Agarwal, V. Balakrishnan, A. Bhuyan, K. Kim, B.C. Paul, W. Wang, B. Yang, Y. Cao, S. Mitra, "Optimized Circuit Failure Prediction for Aging: Practicality and Promise", in Proc. *of IEEE Int. Test Conf*., pp. 1-10, 2008.

[5] S. Borkar, "Electronics Beyond Nano-Scale CMOS", ACM/IEEE Design Automation Conf., 2006.

[6] A. C. Cabe et al., "Small Embeddable NBTI Sensors (SENS) for Tracking On-Chip Performance Decay", in Proc. of Symp. on Quality Electronic Design, pp. 1-6, 2009.

[7] K. Kang, et al., "Characterization and Estimation of Circuit Reliability Degradation Under NBTI Using On-Line IDDQ Measurement", in Proc. of Design Automation Conf., pp. 358-363, 2007.

[8] T. H. Kim et al., "Silicon Odometer: an On-Chip Reliability Monitor for Measuring Frequency Degradation of Digital Circits", IEEE J. Solid State Circuits, Vol. 3, No. 4, pp. 874-880, April 2008.

[9] E. Karl, et al., "Compact In-Situ Sensors for Monitoring Negative-Bias-Temperature-Instability Effect and Oxide Degradation", in Proc. of Solid State Circ. Conf., pp. 410-411, 2008.

[10] C. Thibeault, "On the Comparison of ΔI_{DDQ} and I_{DDQ} Testing", in Proc. of IEEE VLSI Test Symp., pp. 143-150, 1999.

[11] C. Metra, et al. "Self-Checking Detection and Diagnosis of Transient, Delay, and Crosstalk Faults Affecting Bus Lines", IEEE Trans. on Computers, Vol. 49, No. 6, pp.560-574, June 2000.

[12] M. Omaña, D. Rossi, N. Bosio, C. Metra, "Novel Low-Cost Aging Sensor", to appear as a poster in the Proc. of Computing Frontiers, May 17-19, 2010.

[13] D.A. Anderson, G. Metze, "Design of Totally Self-Checking Circuits for m-Out-of-n Codes," *IEEE Trans. Comp.*, Vol. 22, No. 3, pp. 263-269, Mar. 1973.

[14] S. Tam, et al., "Clock Generation and Distribution for the 130-nm Itanium® 2 Processor With 6-MB On-Die L3 Cache", *IEEE J. of Solid-State Circuits*, Vol. 39, No. 4, pp. 636-642, April 2004.

[15] M. Nicolaidis, B. Courtois, "Strongly Code Disjoint Checkers", *IEEE Transactions on Computers*, Vol. 37, No. 6, pp. 751-756, June 1988.

[16] R. Rodriguez-Montanes, E.M.J.G. Bruls, J. Figueras, "Bridging Defect Resistance Measurements in a CMOS Process", in *Proc. of IE EE Int. Test Conf.*, 1992, pp. 892-899.

[17] A. Casimiro, *et al.* , "Experiments on Bridging Fault Analysis and Layout-Level DFT for CMOS Designs", in *Proc. of IEEE Defects and Fault Tol. in VLSI Syst.*, pp. 109-116, 1993.